**Chương 1 : Tổng quan về đề tài**

* 1. **Tổng quan về công nghệ thông tin :**

**1.1.1:Tổng quan về ngành công nghệ thông tin :**

-Ngành công nghệ thông tin (IT – Information Technology) là ngành khoa học kỹ thuật xử dụng máy tính và các phần mềm máy tính để phân phối, xử lý dữ liệu cũng như lưu trữ, trao đổi và sử dụng thông tin. Các sản phẩm và dịch vụ của ngành này đã trở thành một phần không thể thiếu trong cuộc sống hiện đại, từ điện thoại thông minh, máy tính cá nhân cho đến các ứng dụng trực tuyến.

-Tầm quan trọng của ngành công nghệ thông tin trong thế giới hiện đại không thể được phủ nhận. Nó không chỉ cung cấp các giải pháp cho doanh nghiệp, nâng cao hiệu quả sản xuất và quản lý, mà còn cải thiện chất lượng cuộc sống thông qua các ứng dụng trong y tế, giáo dục, và dịch vụ công cộng. Ngành công nghệ thông tin là chìa khóa cho sự đổi mới và phát triển bền vững, mở ra cơ hội nghề nghiệp mới và thúc đẩy tăng trưởng kinh tế.

-Trong thời đại số hóa, ngành công nghệ thông tin ngày càng phát triển mạnh mẽ và tạo ra nhiều cơ hội việc làm cho các lập trình viên. Các “ông lớn” công nghệ như Apple, Google, Amazon, Facebook và Microsoft cũng là những nhà tuyển dụng lớn trong ngành này. Tại Việt Nam, nhu cầu tuyển dụng lập trình viên CNTT luôn ở mức cao.

**1.1.2 : Các nhóm ngành trong lĩnh vực công nghệ thông tin** :

-Lĩnh vực công nghệ thông tin (IT) vô cùng đa dạng và được phân chia thành rất nhiều các nhóm ngành khác nhau, mỗi nhóm tập trung vào một lĩnh vực cụ thể của công nghệ.Dưới đây là một số nhóm ngành chính trong lĩnh vực IT :

**Lập Trình Viên (Software Developer)** : Lập trình viên là một trong những công việc phổ biến nhất đối với sinh viên tốt nghiệp ngành công nghệ thông tin. Lập trình viên có nhiệm vụ thiết kế, phát triển, kiểm thử và duy trì các ứng dụng phần mềm. Các ngôn ngữ lập trình phổ biến mà lập trình viên thường sử dụng bao gồm Java, Python, C#, C++, JavaScript, PHP, Ruby… Lập trình viên có thể làm việc trong nhiều lĩnh vực khác nhau như phát triển ứng dụng di động, phát triển web, phát triển trò chơi điện tử hay phát triển phần mềm doanh nghiệp.

**Chuyên Viên An Ninh Mạng (Cybersecurity Specialist)** : Chuyên viên an ninh mạng có nhiệm vụ bảo vệ hệ thống, dữ liệu khỏi các cuộc tấn công mạng. Công việc của họ bao gồm phát hiện và ngăn chặn các mối đe dọa bảo mật, đánh giá lỗ hổng, thiết lập các chính sách bảo mật và thực hiện các biện pháp phục hồi sau khi bị tấn công. Một trong những yếu tố quan trọng mà họ cần hiểu rõ là hacker là gì, để từ đó có thể xác định các phương pháp và công cụ mà kẻ tấn công có thể sử dụng. Các chuyên viên an ninh mạng cần có kiến thức sâu rộng về các công nghệ bảo mật, mạng máy tính và pháp luật về bảo mật thông tin.

**Quản Trị Hệ Thống (System Administrator**): Quản trị hệ thống có nhiệm vụ quản lý, duy trì các hệ thống máy tính và mạng trong tổ chức. Họ sẽ thực hiện cài đặt, cấu hình, bảo trì và giám sát các máy chủ, mạng và hệ thống lưu trữ dữ liệu. Quản trị hệ thống cũng phải đảm bảo hệ thống hoạt động ổn định, an toàn và hiệu quả, đồng thời hỗ trợ người dùng khi gặp sự cố kỹ thuật.

**Chuyên Viên Phân Tích Dữ Liệu (Data Analyst)** : Chuyên viên phân tích dữ liệu có nhiệm vụ thu thập, xử lý và phân tích dữ liệu để đưa ra các quyết định kinh doanh quan trọng. Vị trí này cần sử dụng các công cụ và kỹ thuật phân tích dữ liệu như SQL, Python, R, Excel, các phần mềm BI (Business Intelligence) để tạo ra các báo cáo, biểu đồ. Chuyên viên phân tích dữ liệu cần có khả năng phân tích, suy luận logic cũng như kỹ năng giao tiếp tốt để trình bày kết quả phân tích cho các bên liên quan.

**Kỹ Sư Mạng (Network Engineer)** : Nhiệm vụ của kỹ sư mạng là thiết kế, triển khai và quản lý các hệ thống mạng máy tính. Người làm vị trí này sẽ có trách nhiệm quản lý các thiết bị mạng như router, switch, firewall và thiết lập các kết nối mạng nội bộ, mạng internet. Kỹ sư mạng cũng cần đảm bảo tính bảo mật, hiệu suất của mạng, khắc phục các sự cố mạng khi xảy ra.

**Chuyên Viên Hỗ Trợ Kỹ Thuật (Technical Support Specialist)** : Chuyên viên hỗ trợ kỹ thuật là vị trí cung cấp hỗ trợ, giải quyết các vấn đề kỹ thuật cho người dùng thông qua việc chẩn đoán, khắc phục các sự cố phần cứng, phần mềm, cài đặt và cấu hình các thiết bị, ứng dụng… Để làm tốt công việc của mình, chuyên viên hỗ trợ kỹ thuật cần có kỹ năng giao tiếp tốt và khả năng giải quyết vấn đề nhanh chóng.

**Quản Trị Cơ Sở Dữ Liệu (Database Administrator) :** Quản trị cơ sở dữ liệu là vị trí đảm nhiệm nhiều công việc quan trọng khác nhau như thiết kế, triển khai, quản lý các hệ thống cơ sở dữ liệu, tối ưu hóa hiệu suất của cơ sở dữ liệu và thực hiện sao lưu và khôi phục dữ liệu khi cần thiết. Quản trị cơ sở dữ liệu cần có kiến thức chuyên sâu về các hệ quản trị cơ sở dữ liệu như MySQL, SQL Server, Oracle và PostgreSQL.

**Chuyên Viên Phát Triển Web (Web Developer)** : Chuyên viên phát triển web là người thiết kế, phát triển các trang web và ứng dụng web. Cụ thể, họ sẽ sử dụng các ngôn ngữ lập trình như HTML, CSS, JavaScript và các framework như Angular, React, và Vue.js để viết mã nguồn. Chuyên viên phát triển web cần có khả năng thiết kế giao diện người dùng thân thiện và tối ưu hóa trải nghiệm người dùng trên các thiết bị khác nhau.

* 1. **Tổng quan về phân tích dữ liệu :**

**1.2.1:Phân tích dữ liệu là gì :**

- **Phân tích dữ liệu**là môn khoa học phân tích dữ liệu thô (raw data) để đưa ra được kết luận về thông tin đó. Các chuyên gia Phân tích dữ liệu tìm ra các xu hướng và số liệu trong các khối thông tin mà có thể bị bỏ sót nếu không sử dụng kĩ thuật hay công cụ phân tích. Thông tin thu được có thể được sử dụng để tối ưu hóa các quy trình làm tăng hiệu quả tổng thể của một doanh nghiệp hoặc một hệ thống

-**Data Analyst** là **công việc tập trung vào khai thác, xử lý bộ dữ liệu để đưa ra báo cáo về một vấn đề cụ thể.** Mục đích của một báo cáo phân tích dữ liệu là để giúp nắm được tình hình thực tế và để tham mưu đưa ra quyết định đúng đắn.

**-Đây là một công việc có tầm quan trọng lớn đối với bất cứ tổ chức nào.** Để có thể hiểu được khánh hàng muốn gì, các tổ chức và doanh nghiệp cần người phân tích dữ liệu đã thu thập từ khách hàng

**1.2.2:Quy trình phân tích dữ liệu :**

**-** **Thu thập dữ liệu :**  là bước đầu tiên giúp doanh nghiệp xác định nguồn tài nguyên dữ liệu sẵn có hiện tại. Một số hoạt động diễn ra trong bước này bao gồm :

+ **Xác định mục tiêu**: là những kết quả cụ thể mà ta muốn đạt được thông qua việc xử lý và phân tích dữ liệu. Mục tiêu này xác định hướng đi và phạm vi của quá trình phân tích, giúp ta tập trung vào việc thu thập thông tin quan trọng và thực hiện các phân để đáp ứng các yêu cầu hoặc nhu cầu cụ thể.  Phân tích dữ liệu phục vụ mục đích gì? Bạn cần kết quả nào ở việc phân tích? Bạn chỉ cần liệt kê và trả lời những câu hỏi cần thiết để xác định mục tiêu của mình.

+ **Thu thập dữ liệu**: là thu thập dữ liệu từ các nguồn khác nhau như cơ sở dữ liệu, tệp tin, trang web, thiết bị cảm biến, và nhiều nguồn khác. Dữ liệu có thể là số liệu, văn bản, hình ảnh, hoặc âm thanh.

**-Tiền xử lý dữ liệu :** là quá trình chuẩn bị dữ liệu để chuẩn bị cho các bước phân tích hoặc mô hình hóa sau này.  Một số hoạt động quan trọng có thể kể tới:

**+ Làm sạch dữ liệu**: Loại bỏ các dữ liệu sai, không hợp lệ, trùng lặp hoặc bị nhiễu; đồng thời dữ liệu sẽ được kiểm tra, sửa các giá trị lỗi và chuyển đổi định dạng về cùng một đơn vị hoặc thang đo nếu cần.

**+ Xử lý dữ liệu thiếu:** Xác định và xử lý các dữ liệu bị thiếu trong tập dữ liệu bằng một số phương pháp như: điền giá trị trung bình, giá trị gần nhất, hoặc dự báo để điền dữ liệu có độ sai số ít nhất. Hoặc trong nhiều trường hợp, các dòng hoặc cột dữ liệu có quá nhiều giá trị thiếu không thể khôi phục sẽ được loại bỏ.

**- Phân tích dữ liệu:** Đây là một bước vô cùng quan trọng trong quá trình nghiên cứu và đưa ra quyết định, bởi nó giúp chúng ta khai thác giá trị tiềm ẩn từ những dữ liệu sẵn có.Tùy thuộc vào mục tiêu và đặc điểm của dữ liệu, quá trình phân tích có thể sử dụng nhiều phương pháp khác nhau, từ các kỹ thuật phân tích mô tả cơ bản, phân tích hồi quy để tìm mối quan hệ giữa các biến, phân tích sự khác biệt để so sánh các nhóm dữ liệu, cho đến các phương pháp thống kê nâng cao giúp đưa ra những kết luận chính xác và có cơ sở. Bên cạnh đó, các kỹ thuật hiện đại như machine learning (học máy) và data mining (khai thác dữ liệu) cũng được ứng dụng rộng rãi để xử lý lượng dữ liệu lớn và phức tạp, phát hiện ra những mẫu và dự đoán các xu hướng tiềm năng.

-**Kết luận và dự đoán**: Dựa trên việc phân tích kỹ lưỡng các dữ liệu hiện có và thông tin thu thập được từ các nguồn đáng tin cậy, chúng ta có thể rút ra những kết luận quan trọng, giúp chúng ta hiểu rõ hơn về tình hình hiện tại cũng như các yếu tố tác động đến sự phát triển của vấn đề đang được xem xét. Việc phân tích này không chỉ cung cấp cái nhìn sâu sắc về thực trạng, mà còn giúp chúng ta nhận diện các xu hướng và mô hình có thể xảy ra trong tương lai. Từ đó, chúng ta có thể đưa ra những dự đoán hợp lý về các sự kiện sắp tới, từ đó đề xuất các giải pháp hoặc chiến lược phù hợp để ứng phó và tận dụng cơ hội, đồng thời giảm thiểu các rủi ro có thể xảy ra.

**1.2.3:Một số công cụ phân tích dữ liệu phổ biến :**

-**Excel** :là một phần mềm được phát triển bởi Microsoft, cung cấp đa dạng tính năng để phân tích dữ liệu trong các bảng tính: thực hiện các phép tính, sắp xếp dữ liệu, lọc dữ liệu, tạo biểu đồ, và thậm chí áp dụng các công thức phức tạp để xử lý thông tin. Ngoài ra, Excel cũng hỗ trợ các công cụ như PivotTables và PivotCharts để hỗ trợ trực quan hoá dữ liệu.

Excel có ưu điểm là dễ cài đặt, giao diện dễ sử dụng, và gần như đã phổ biến trong tất cả các doanh nghiệp hiện nay. Tuy nhiên, đối với các tập dữ liệu lớn và phức tạp, Excel chưa đủ mạnh để xử lý như các công cụ Business Intelligence chuyên sâu.

-**Power PI**là một công cụ phân tích và trực quan hóa dữ liệu phổ biến được phát triển bởi Microsoft, với các tính năng mạnh mẽ phục vụ hoạt động thu thập, làm sạch, xử lý và trực quan hóa dữ liệu cho người dùng.

Power BI cho phép tạo ra các biểu đồ, bảng biểu, bảng điều khiển,… để thể hiện thông tin từ dữ liệu. Đồng thời, tính năng chia sẻ dữ liệu dễ dàng cũng hỗ trợ đắc lực cho doanh nghiệp trong hoạt động cộng tác nội bộ, và ra quyết định data-driven.

-**Python** là một trong những ngôn ngữ lập trình dễ nhất, linh hoạt và có tính ứng dụng cao. Chúng cung cấp các thư viện và framework phổ biến như NumPy, Pandas hay Matplotlib có khả năng làm sạch, chuyển đổi, phân tích và trực quan hoá hệ thống dữ liệu.

Nếu NumPy cung cấp mảng và tính toán số học nhanh chóng thì Pandas đưa ra các cấu trúc dữ liệu linh hoạt đi kèm với công cụ phân tích bổ trợ khác nhau. Ngoài ra, Matplotlib tạo biểu đồ và đồ thị để trực quan hóa dữ liệu. Khi đó, bộ ba này là công cụ đắc lực cho việc phân tích phức tạp và xây dựng mô hình dự đoán dữ liệu tương lai.

**1.3**.**Bài toàn phân tích dự đoán sự phân phối dân số dựa trên mô hình hồi quy tuyến tính.**

**1.3.1.Tổng quan về phân tích dự báo:**

**-Khái niệm :** Phân tích dự báo là việc sử dụng dữ liệu kết hợp với các thuật toán thống kê, kỹ thuật trong “Machine learning” để xác định những khả năng, kết quả có thể xảy ra của đối tượng nghiên cứu trong tương lai dựa trên dữ liệu lịch sử. Mục tiêu của phân tích dự báo là vượt lên trên các hiểu biết về những gì đã xảy ra ở quá khứ để đánh giá tốt nhất những gì sẽ xảy ra trong tương lai.

**1.3.1.1.Lịch sử của phân tích dự báo :** Lịch sử bài toán dự đoán gắn liền với sự phát triển của thống kê, máy tính và các lĩnh vực như học máy, trí tuệ nhân tạo, và khoa học dữ liệu. Dưới đây là một cái nhìn tổng quan về lịch sử của bài toán dự đoán:

- Thống kê cổ điển (Ancient Statistics): Ý tưởng về dự đoán và ước lượng có nguồn gốc từ thống kê cổ điển, nơi mà những phương pháp đơn giản như trung bình, trung vị và độ lệch chuẩn đã được sử dụng để mô tả và dự đoán dữ liệu.

**- Phân tích hồi quy (Thế kỷ 17 - 18):** Phương pháp hồi quy xuất hiện trong thế kỷ 17 và 18 với công trình của nhà toán học John Graunt và nhà thống kê Francis Galton. Họ đã phát triển các kỹ thuật để mô tả mối quan hệ giữa các biến và dự đoán giá trị tương lai.

**- Phát triển máy tính (Thế kỷ 20):** Sự phát triển của máy tính trong thế kỷ 20 đã mở ra những khả năng mới cho bài toán dự đoán. Các phương pháp tính toán, bao gồm cả máy tính analog và digital, đã giúp tăng cường khả năng tính toán và xử lý dữ liệu.

**- Học máy (Thế kỷ 20 – Đầu thế kỉ 21):** Sự xuất hiện của máy học vào cuối thế kỷ 20 và đầu thế kỷ 21 đã làm thay đổi cách chúng ta tiếp cận bài toán dự đoán. Các thuật toán như cây quyết định, máy vector hỗ trợ, và mạng nơ-ron đã trở thành những công cụ phổ biến để xây dựng các mô hình dự đoán phức tạp.

**- Big Data và Khoa học Dữ liệu (Thế kỷ 21):** Sự gia tăng về kích thước và đa dạng của dữ liệu, cùng với sự phát triển của công nghệ, đã làm tăng cường khả năng xử lý dữ liệu và xây dựng các mô hình dự đoán. Khoa học dữ liệu đã trở thành một lĩnh vực chính trong nghiên cứu và công nghiệp.

**- Deep Learning (2010s - hiện tại):** Một trong những phát triển quan trọng nhất là sự nổi lên của deep learning, đặc biệt là mạng nơ-ron sâu (deep neural networks). Điều này đã mang lại sự tiến bộ đáng kể trong khả năng mô hình hóa và dự đoán, đặc biệt là trong các ứng dụng như thị giác máy tính và xử lý ngôn ngữ tự nhiên.

**- Dự báo trong thời đại số hóa (Hiện nay):** Ngày nay, bài toán dự đoán trở thành một phần quan trọng trong nhiều lĩnh vực, từ kinh doanh, tài chính, y tế đến giáo dục và nhiều lĩnh vực khác. Công ty và tổ chức sử dụng mô hình dự đoán để đưa ra quyết định thông minh và hiệu quả.

Bài toán dự báo là một trong những thách thức quan trọng trong lĩnh vực phân tích dữ liệu, nơi chúng ta cố gắng dự đoán giá trị của một biến mục tiêu trong tương lai dựa trên dữ liệu lịch sử và các yếu tố ảnh hưởng. Mục tiêu chính của bài toán dự báo là xây dựng một mô hình có khả năng hiểu và ứng dụng các mẫu, xu hướng và quy luật từ dữ liệu để thực hiện việc dự đoán một cách chính xác và đáng tin cậy.

**1.3.1.2.Lợi ích của phân tích dự báo :**

**-** **Phân tích dự báo** thường được một công ty hay tổ chức sử dụng mục đích giúp giải quyết các vấn đề khó khăn, hạn chế rủi ro hay tìm ra được điểm mạnh, điểm yếu của mình để cải thiện hoạt động, và khám phá những cơ hội kinh doanh mới trong tương lai.

* **Phát hiện, ngăn chặn các hành vi tội phạm, lừa đảo**: bằng cách kết hợp nhiều phương pháp phân tích bao gồm phân tích dự báo, công ty có thể cải thiện khả năng phát hiện ngăn chặn hành vi tội phạm. Khi an ninh mạng trở thành mối quan tâm đặc biệt ngày nay, các công cụ phân tích hành vi tối ưu sẽ có khả năng kiểm tra tất cả các hành động trên mạng máy tính trong thời gian thực qua đó phát hiện những vấn đề bất thường có thể xuất phát từ hành vi của bọn tội phạm, gian lận.
* **Tối ưu hóa các chiến dịch Marketing:** phân tích dự báo được sử dụng để xác định các nhóm khách hàng tiềm năng, dự báo xu hướng tiêu dùng, hành vi, khả năng mua hàng, nhu cầu, mong muốn của khách hàng, hay sản phẩm khách hàng ưa thích nhất trong tương lai qua đó giúp công ty phát triển các chiến dịch tiếp thị hiệu quả phù hợp với nhu cầu khách hàng, có thể tăng trải nghiệm khách hàng, giữ chân họ lâu hơn.
* **Cải thiện hiệu quả hoạt động, sản xuất:** nhiều công ty sử dụng các mô hình dự báo để dự báo hàng tồn kho và quản lý tài sản, nguồn lực phục vụ cho quá trình kinh doanh. Ví dụ các hãng hàng không sử dụng phân tích dự báo để thiếp lập giá vé. Các khách sạn cố gắng dự báo số lượng khách cho mỗi đêm để tối đa hóa phân phối, sắp xếp phòng trống cho khách hàng và tăng doanh thu. Các công ty logistics và vận tải thường sử dụng các phân tích dự báo để cải thiện việc thiết lập thời gian biểu và lịch trình, cũng như đảm bảo cho các phương tiện hoạt động ở hiệu suất cao nhất.
* **Phát triển các cơ hội kinh doanh:** phân tích dự báo đặc biệt có giá trị với các công ty khi xác định cơ hội kinh doanh mới. Các công ty có thể sử dụng phân tích dự báo để đánh giá tiềm năng thị trường cho các sản phẩm mới, dự báo nhu cầu hàng tồn kho để làm sao lượng cung ứng sản phẩm ra thị trường đạt mức tối ưu.
* **Giảm thiểu rủi ro:** ví dụ trong ngành tài chính, credit score, điểm tín dụng được sử dụng để đánh giá khả năng thanh toán các khoản tín dụng của khách hàng, chính là một ví dụ điển hình của phân tích dự báo. Điểm tín dụng được tạo bởi mô hình dự báo kết hợp tất cả dữ liệu liên quan có thể sử dụng để phân tích, đánh giá mức độ uy tín của khách hàng. Qua các điểm tín dụng được tính toán, công ty, tổ chức tài chính sẽ biết được trong tương lai khách hàng nào có nguy cơ không thể thanh toán khoản nợ, đưa ra giải pháp phù hợp để giảm thiểu rủi ro.
* **Giảm chi phí:** rủi ro thấp hơn sẽ dẫn đến chi phí thấp hơn. Nếu công ty càng có thể dự đoán tương lai tốt hơn sẽ có nhiều bước đi tốt hơn để đảm bảo rằng các tác động tiêu cực hay rủi ro tài chính sẽ được hạn chế tối đa. Ví dụ: nhiều công ty sử dụng phân tích dự báo để xác định nhu cầu bảo trì các máy móc thiết bị trước khi xảy ra sự cố, kiểm soát thực trạng của chúng trong khi chúng vẫn đang hoạt động. Đây là cách thức chủ động giải quyết và ngăn ngừa sự cố, tăng hiệu suất và giảm chi phí liên quan đến hư hỏng máy móc thiết bị, chậm tiến độ vận hành các quy trình sản xuất, kinh doanh khác.
* **Gia tăng lợi thế cạnh tranh:** thông qua phân tích dự báo, một công ty có thể không chỉ dự báo được hành vi khách hàng, nhu cầu khách hàng mà còn dự báo được các yếu tố khác từ thị trường kinh doanh có thể tác động lên lợi nhuận của họ. Những thông tin công ty có được từ phân tích dự báo sẽ giúp đề xuất những chiến lược cạnh tranh hiệu quả trước các đối thủ.

Phân tích dự báo là cơ sở để các công ty có thể tiếp cận với nhiều nền tảng công nghệ mới, hỗ trợ cho các dự án khai thác dữ liệu. Khi công nghệ Machine learning, AI, IoT được ứng dụng phổ biến hơn, thì các công cụ phân tích dự báo sẽ được phát triển tối ưu hơn. Ngày nay ở các công ty quy mô lớn, phân tích dự báo được sử dụng kết hợp với Machine learning để xử lý dữ liệu tự động, nhanh hơn, hiệu quả hơn.

**1.3.1.3. Ứng dụng của phân tích dự báo trong một số ngành và lĩnh vực :**

* **Lĩnh vực tài chính, ngân hàng :**Trong ngành tài chính, với khối lượng dữ liệu và số tiền giao dịch là khổng lồ, và luôn bị đe dọa. Các tổ chức hoạt động trong lĩnh vực này từ lâu đã triển khai các công cụ phân tích dự báo để phát hiện và giảm thiểu hành vi gian lận, tội phạm, đo lường rủi ro tín dụng, tối đa hóa cơ hội cung cấp các dịch vụ cao hơn (cross-sell, up-sell) và giữ chân khách hàng, đem lại lợi nhuận. Ngân hàng Commonwealth của Úc sử dụng các phân tích để dự đoán khả năng xuất hiện hành vi bất thường cho bất kỳ giao dịch cụ thể nào trước khi được thực hiện – trong vòng 40 mili giây kể từ khi bắt đầu giao dịch. Ngoài ra, các tổ chức còn sử dụng phân tích dự báo để dự báo xu hướng thị trường tài chính, dự đoán các tác động của các chính sách, luật và quy định mới của quốc gia.
* **Lĩnh vực E-commerce và bán lẻ :** Các nhà bán lẻ ở ngày nay đang sử dụng các phân tích dự báo để lập kế hoạch bán hàng và tối ưu hóa giá cả, để phân tích hiệu quả của các sự kiện khuyến mãi và xác định ưu đãi nào phù hợp nhất với người tiêu dùng. Cụ thể mục đích triển khai phân tích dự báo ở các công ty thương mại điện tử, và bán lẻ bao gồm: cải thiện trải nghiệm khách hàng, giảm tỷ lệ khách hàng rời dịch vụ, xây dựng hệ thống giá cả linh hoạt, marketing theo hướng cá nhân hóa, cải thiện khả năng quản lý hàng tồn kho, cải thiện việc phân bổ vị trí các cửa hàng, thiết lập các kênh bán hàng sao cho phù hơp.
* **Công nghiệp năng lượng :** Các tổ chức, công ty hoạt động lĩnh vực năng lượng sử dụng phân tích dự báo để dự báo lỗi, máy móc thiết bị và nhu cầu tài nguyên trong tương lai, giảm thiểu các rủi ro về độ an toàn, cải thiện hiệu suất hoạt động tổng thể. Cụ thể nó được sử dụng để dự đoán nhu cầu và nguồn cung về năng lượng, bằng cách các mô hình phức tạp nghiên cứu công suất của các nhà máy, tác động của sự thay đổi thời tiết và các yếu tố khác như thời vụ hàng năm,.. giúp các tổ chức trong ngành năng lượng có thể tiết kiệm các nguồn tài nguyên quý giá.
* **Tổ chức hoạt động trong lĩnh vực công :** Theo SAS, các tổ chức Chính phủ, cơ quan Nhà nước đã đóng vai trò quan trọng trong sự tiến bộ của công nghệ máy tính. Cục điều tra dân số Hoa Kỳ, và ở các nước khác đã phân tích dữ liệu để nắm bắt xu hướng dân số trong nhiều thập kỷ. Các tổ chức Chính phủ ngày nay sử dụng phân tích dự báo như nhiều ngành công nghiệp khác – để cải thiện các dịch vụ công và hiệu suất làm việc; phát hiện và ngăn chặn tội phạm; và hiểu rõ hơn hành vi của người tiêu dùng. Họ cũng sử dụng các phân tích dự báo để tăng cường an ninh mạng.
* **Lĩnh vực y tế, sức khỏe :** Theo Wikipedia, các chuyên gia sử dụng phân tích dự báo trong lĩnh vực chăm sóc sức khỏe chủ yếu để xác định bệnh nhân nào có nguy cơ mắc một số bệnh nghiêm trọng. Các chuyên gia chăm sóc sức khỏe có thể sử dụng các phân tích dự báo để phân tích dữ liệu bệnh nhân, có thể giúp các bác sĩ dự báo khả năng mắc bệnh và giúp chẩn đoán sớm. Các biện pháp dự đoán như thế có thể làm giảm việc bệnh nhân tái nhập viện, giảm chi phí chăm sóc sức khỏe, xác định bệnh nhân nào có tình trạng xấu, nguy cơ tử vong cao nên được ưu tiên, giảm thời gian chờ đợi ở bệnh viện,…Đặc biệt là hệ thống “Clinical decision support” được xây dựng dựa trên phân tích dự báo, có khả năng cung cấp cho các bác sĩ, nhân viên y tế, bệnh nhân các kiến thức cần thiết và thông tin cụ thể về người bệnh, nhằm hỗ trợ chăm sóc sức khỏe cho người bệnh khi cần thiết. Nó bao gồm một loạt các công cụ và tính năng như cảnh báo tự động và nhắc nhở người bệnh, bác sĩ, nhân viên y tế, cung cấp hướng dẫn lâm sàng, hỗ trợ chẩn đoán,…
* **Lĩnh vực sản xuất :** Đối với các nhà sản xuất, việc xác định các yếu tố, nguyên nhân dẫn đến giảm chất lượng và thất bại trong sản xuất rất quan trọng. Các nhà sản xuất sử dụng phân tích dự báo để lên kế hoạch bảo trì máy móc thiết bị, có thể lường trước sự cố của thiết bị, dự báo nhu cầu năng lượng và giảm chi phí vận hành, sửa chữa. Lenovo chỉ là một nhà sản xuất đã sử dụng các phân tích dự báo để hiểu rõ hơn về nhu cầu bảo hành xe – một sáng kiến giúp giảm 10 đến 15% chi phí bảo hành. Ví dụ, các máy cảm biến đo độ rung trong các bộ phận ô tô có thể báo hiệu sự cần thiết phải bảo trì trước khi xe bị hỏng trên đang di chuyển. Trong một công ty kinh doanh, việc sử dụng phân tích dự báo có thể đem lại lợi ích cho các nhà máy sản xuất ví dụ như dự báo nhu cầu sản phẩm, lượng sản phẩm có thể sản xuất để dự báo nguồn nguyên vật liệu, công suất khả thi của máy móc, thiết bị, lên kế hoạch phân bổ nguồn lược, cải thiện quy trình sản xuất, tối ưu việc sử dụng máy móc, thiết bị, dự báo máy móc thiết bị nào có thể bị hỏng hóc, lên kế hoạch bảo trì, v.v

**1.3.1.4:Thách thức và giải pháp của phân tích dự báo :**

**- Khối lượng vượt quá khả năng quản lý :** Các công ty hiện đang sở hữu hàng terabyte, thậm chí hàng exabyte dữ liệu. Con số này không ngừng tăng lên. Do đó, dữ liệu có thể dễ dàng vượt khỏi tầm kiểm soát nếu không được quản lý đúng cách, đồng thời, doanh nghiệp sẽ bỏ lỡ cơ hội khai thác giá trị từ tài sản dữ liệu.

**Giải pháp :**

Sử dụng công nghệ quản lý và lưu trữ để giải quyết khối lượng ngày càng tăng và thách thức trong việc quản lý Big data. Cho dù là đám mây lưu trữ tại chỗ hay phương pháp kết hợp, hãy đảm bảo lựa chọn đó phù hợp với mục tiêu kinh doanh và nhu cầu của tổ chức.

Tạo một kiến ​​trúc có khả năng mở rộng với những công cụ có thể điều chỉnh theo khối lượng dữ liệu mà không ảnh hưởng đến tính toàn vẹn của nó.

- **Dữ liệu kém dẫn đến kết quả không như mong muốn** : Chất lượng kém là một trong những thách thức lớn nhất của Big data. Dữ liệu chất lượng kém dẫn đến sai sót trong thông tin thu lại và hiểu biết của người sử dụng dữ liệu bị sai lệch. Cuối cùng, doanh nghiệp phải gánh chịu hậu quả bằng cách tăng chi phí kinh doanh và hơn thế nữa.

**Giải pháp :**

Bước đầu tiên để đảm bảo dọn dẹp dữ liệu “sạch sẽ” là có quy trình và nhân sự chăm sóc dữ liệu chuyên nghiệp trong tổ chức. Doanh nghiệp cần thiết lập quy trình quản trị dữ liệu đầy đủ để xác định các công cụ, hoạt động quản lý dữ liệu và kiểm soát truy cập.

Điều cần thiết là phải hiểu rõ về cách sắp xếp và bóc tách dữ liệu dựa trên mục tiêu kinh doanh. Vì vậy, doanh nghiệp có thể sẽ cần mời những chuyên gia về dữ liệu hoặc nhân viên kinh doanh, người thực sự sử dụng dữ liệu này để xác định yêu cầu về chất lượng dữ liệu.

- **Xử lý nhiều định dạng :** Trong thực tế, hầu hết dữ liệu tổ chức, doanh nghiệp hoặc cá nhân thu thập được đều không có cấu trúc hoặc bán cấu trúc (ngoại trừ trường hợp dữ liệu nằm trong bảng cơ sở dữ liệu như email, đánh giá của khách hàng, video,...).

Điều này đưa đến một loạt thách thức, đó là tìm ra cách đưa dữ liệu không đồng nhất sang định dạng phù hợp với nhu cầu kinh doanh, đồng thời, phù hợp với yêu cầu của các công cụ doanh nghiệp đang sử dụng trong quá trình phân tích thông tin, trực quan hóa , dự đoán,...

**Giải pháp :**

Tìm hiểu cách sử dụng công cụ và công nghệ xử lý dữ liệu hiện đại để định dạng lại dữ liệu phi cấu trúc và rút ra hiểu biết sâu sắc từ đó. Nếu xử lý nhiều định dạng, có thể kết hợp nhiều công cụ khác nhau để phân tích cú pháp dữ liệu và trích xuất thông tin cần thiết.

Áp dụng hoặc tạo ra ứng dụng tùy chỉnh nhằm tăng tốc và tự động hóa quá trình chuyển đổi dữ liệu thô thành thông tin chi tiết, có giá trị. Việc lựa chọn công cụ sẽ tùy thuộc vào nguồn và bản chất của dữ liệu, cũng như yêu cầu riêng của doanh nghiệp.

- **Chi phí cao cho dự án và cơ sở hạ tầng dữ liệu :** Theo một số nghiên cứu, 50% giám đốc điều hành ở Mỹ và 39% giám đốc điều hành ở châu Âu Thừa nhận rằng ngân sách công nghệ thông tin hạn chế là một trong những rào cản lớn nhất ngăn họ tận dụng dữ liệu của mình.

Việc triển khai dữ liệu lớn rất tốn kém, đòi hỏi phải lập kế hoạch cẩn thận và khả năng đáp ứng khoản chi phí đáng kể, nhưng chưa chắc đã nhận lại kết quả tích cực ngay lập tức.

Hơn nữa, khi khối lượng dữ liệu tăng theo cấp số nhân, thì cơ sở hạ tầng cũng vậy. Trên thực tế, có tới 30% số tiền chi trả cho đám mây bị lãng phí, theo *Flexera*.

**Giải pháp :**

Doanh nghiệp có thể giải quyết hầu hết vấn đề chi phí bằng cách liên tục giám sát cơ sở hạ tầng. Có nhiều phương pháp theo dõi dịch vụ và tài nguyên đang sử dụng cho mục đích lưu trữ và quản lý dữ liệu, xác định cơ hội tiết kiệm và cân bằng chi phí.

Cân nhắc chi phí sớm trước khi tạo quy trình xử lý dữ liệu. Dữ liệu trùng lặp, chưa được phân loại, thất lạc,... có thể làm tăng chi phí. Doanh nghiệp cần nghiên cứu tài nguyên dữ liệu hiện có để đưa ra chiến lược phù hợp và tối ưu hóa chi phí.

Hãy chọn công cụ tiết kiệm chi phí phù hợp với ngân sách. Hầu hết dịch vụ đám mây đều được cung cấp dựa trên cơ sở chi trả theo thực tế sử dụng, nghĩa là chi phí phụ thuộc trực tiếp vào các dịch vụ và sức mạnh tính toán mà doanh nghiệp sử dụng.

**- Chưa tận dụng tối đa thông tin chi tiết :** Trích xuất thông tin chi tiết là một chuyện. Đưa chúng vào sử dụng là một câu chuyện hoàn toàn khác. Nếu bước thứ hai không hoạt động, toàn bộ chiến lược dữ liệu lớn của tổ chức có thể thất bại vì nó không thể mang lại bất kỳ nguồn lợi nhuận nào.

**Giải pháp :**

Đặt ra những tình huống kinh doanh cụ thể để nhân viên kinh doanh hiểu rõ hơn họ cần gì từ dữ liệu và cách họ có thể hành động trên dữ liệu đó.

Sử dụng phân tích nâng cao giúp khám phá thêm phương thức mới để đọc và hiểu thông tin chuyên sâu, đồng thời cung cấp thông tin chi tiết cho bất kỳ ai trong tổ chức.

Cung cấp công cụ trực quan hiện đại, bảng điều khiển, trải nghiệm tương tác và giao diện trực quan để khai thác dữ liệu, khám phá thông tin chuyên sâu, tạo báo cáo và trao đổi thông tin trong tổ chức.